STANFORD UNIVERSITY

Identifying Negative Influencers in Mobile Customer Churn

Manojit Nandi
Verizon Wireless
December 10, 2014

1 INTRODUCTION

Customer churn, the loss of customers for a company, is one of the biggest loss of revenue for Verizon Wireless and other wireless telecommunications companies. Many companies use the lifetime value of a customer, the expected future cash flow from a particular customer, to determine long-term allocation of resources and human capital. As a result, the sudden defection of customers can greatly hinder the long-term growth of these companies. For most companies, providing discounts and other incentives to stop a customer from churning is relatively less expensive than creating a subscription policy with a new customer, so companies prefer to retain customers rather than replace them. Historically, business analysts have used logistic regression models and decision trees to identify which customers are at high-risk for churning.

However over the last decade, companies have looked for ways to improve on these base churn models. Statistical models often assume independence of observation, but this assumption may not be realistic. Individuals who change their wireless carriers can convince their acquaintances to change carriers as well. These social interaction effects are not captured by traditional churn models, so analysts have been looking towards social network methods to construct predictive churn models that leverage social interactions to better identfy potential churners.

1.1 Project Summary

There are two goals for this project. First, I want to develop a social network model that identifies individuals who negatively influence other individuals in the network to churn after them and identify demographic features that could be used to quickly predict negative influencers. Second, I want to compare predictive churn models that incorporate social network metrics as features against Verizon's current churn models. If these new churn models outperform our current models, then we may permenently switch to these social network churn models.

2 Prior Work

Mohatari et. al[1] outline a framework for identifying individuals who influence others to churn in a wireless subscriber network. Their methodology uses the ego-network of each individual who churned to compute a negative influence score. For each node $\underline{u}$ they define three parameters:

1. The neighbors of $\underline{u}$ who subsequently churn within some time period T after $\underline{u}$ churns.
2. The neighbors of $\underline{u}$ who do not churn within some time period T after $\underline{u}$ churns.
3. The prior churn probability p_i for each node i in the network.

For each node $\underline{u}$, the influence score is calculated as follows:

$$Influence(\underline{u}) = \sum_{v' \in Churned(u)} (1 - p_v) - \sum_{v \in NotChurned(v)} p_v \tag{2.1}$$

This influence metric rewards $\underline{u}$ for convincing neighbors with low prior churn probabilities to churn and penalizes $\underline{u}$ for failing to convince neighbors with high prior churn probabilities to churn. After an influence score is assigned to each node in the network, the authors perform regressions with the influence score as the target variable to identify demographic factors that are highly predictive of larger influence scores. This is because the network-based calculation cannot be performed quickly enough for telecommunication companies to prevent churn.

While this paper provides a framework for identifying negative influencers in the network, the authors did not utilize the negative influence scores to predict future churn on an individual level. Kusuma et. al[2] developed various models that incorporate social network analysis to perform customer churn prediction. One of the results these authors found is that incorporate both customer attributes and social network statistics outperform models that use only customer attributes. Building on top of this result, I will incorporate the negative influence scores into Verizon's customer attribute based churn models.

3 Algorithms and Methodology

For this project, I define the communication network $G = (V, E)$ as an unweighted directed graph where there exists an edge e_{ij} between individuals i and j if individual i has called individual j in the given time period T. For each node v in the network, I have a prior probability of churn $\hat{p}_v$ which reflects the probability that user v would have churned with no social influence.

For each node v that churned during the time period T, I compute the negative influence of v, $NegInfluence(v)$, using the following equation:

$$NegInfluence(v) = \sum_{v' \in Churned(v)} \log_2(\frac{1}{\hat{p}_{v'}}) - \sum_{v' \in NotChurned(v)} \log_2(\frac{1}{1-\hat{p}_{v'}}) \quad (3.1)$$

Here, $Churned(v)$ is the set of neighbors of v that churned subsequently after v churned, and $NotChurned(v)$ is the set of neighbors of v that did not churn after v churned. The weights within each summation correspond to the information gain of the event occuring. The quantities $\log_2(\frac{1}{\hat{p}_{v'}})$ places more weight on churn events where the user's prior churn probability is low (closer to 0).For any node u that did not churn during the time period T, the $NegInfluence(u) = 0$ as the sets $Churned(u)$ and $NotChurned(u)$ are both empty sets by construction.

As I explained in the introduction, companies calculate the lifetime value of a customer as the expected revenue stream of a particular customer, and customers with lower prior churn probabilities have higher lifetime value as we believe they will remain customers for a long time and we have to do little to keep them as customers, so we generate a higher revenue stream from them. Therefore, the sudden churn of these customers reflects a greater loss in our projected revenue stream, so this weighting reflects we would prefer to retain these individuals as customers.

To identify factors that are highly predict of high negative influence score, I ran a linear regression where the negative influence score of a node is the response, and various demographic and phone usage variables, such as gender, total minutes used, and total data used.

4 Data Collection and Network Construction

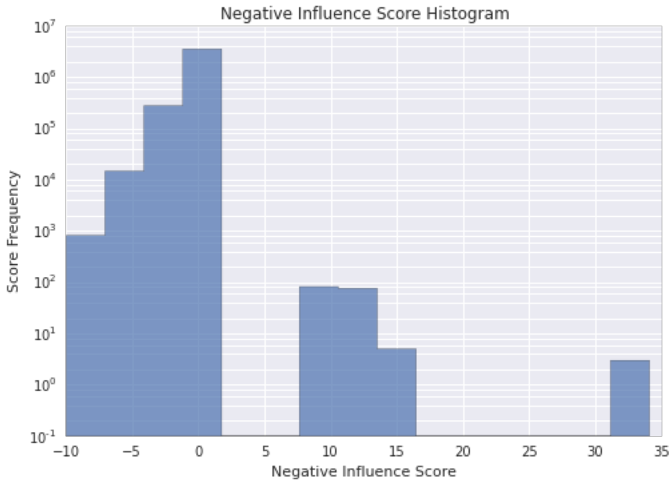
For this project, I pulled the metadata about calls captured by Verizon's towers between September 1, 2014 and September 30, 2014 from Verizon's Call Detail Records (CDR). For each phone call captured, the metadata contains an unique twelve alphanumeric identifier, the date the phone call occured, the time the phone call occurred, the length of the phone call, the phone number of the individual who initiated the call, and the phone call of the recipient.

Verizon's towers capture hundreds of millions of phone calls each day, so to avoid working with a billion node network, I employed breadth-first search to build the network. I selected the root node uniformly at random from a set of Verizon customers who called at least 10 different phone numbers during the month of September. I need the prior churn probabilities and demographic information of each individual in the network, so the beadth-first search process added nodes to the network only if the node's phone number is associated with a Verizon customer. The breadth-first search ran for 16 hops and generated a network with 3,841,197 nodes and 8,922,551 edges.

5 Results

5.1 Negative Influence

For most of the nodes in the network, the negative influence score was 0. The highest negative influence score observed is 34.77, the lowest negative influence score observed is -9.64, and the mean negative influence score is -0.17. A histogram of the negative influence scores are provided below.



From the histogram, we see most of the negative influence scores are either 0 or a negative value. The 0 scores correspond to individuals who did not churn during the one month period, and the negative values correspond to individuals for whom the weights of the non-churning neighbors exceeded the weights of the neighbors who subsequently churned. Furthermore, we observe one outlier corresponding to the node with the highest influence score.

Next, I ran a linear regression using 83 features corresponding to demographic information or wireless usage information to predict the negative influence score. Since there are too many features to list, I have listed the regression summary for the variables with the five highest regression coefficients below to show a subset of the significant predictors.

Variable	Coeff	St.Error	t-Stat	p-value
TOT_MOU_PK	0.744	0.03	24.7	*0.000
URBAN_2	0.561	0.03	18.6	*0.000
SMS_CNT	0.368	0.017	15.8	*0.000
DATA_USG	0.323	0.008	27.9	*0.000
AVG_DUR	0.184	0.02	9.2	*0.000

The first predictor TOT_MOU_PK is the customer's total monthly peak minutes, or the total amount the customer talked with someone else during weekdays. This means a customer who talked more during weekdays is more likely to have a higher influence score.
The second predictor $URBAN\_2$ is an indicator variable meaning this customer lives in a suburban area. The third predictor SMS_CNT is the number of SMS (text) messages the user sent. The fourth predictor $DATA\_USG$ is the amount data the user used for the month, measured in gigabytes, and the last predictor listed AVG_DUR is the average duration of all phone calls for the month, measured in seconds.

5.2 Churn Prediction

To determine whether the social network provides additional predictive power in predicting churn, I used Verizon's current churn prediction model as a base model and added additional features corresponding to social network statistics to create two additional models. In the first model, $SIMPLE\_NET$ model, I added each user's in-degree and out-degree as features. The in-degree and the out-degree correspond to network statistics that can be computed quickly, so if these two statistics show signs of improving churn prediction, then we can easily compute these for larger datasets. For the ADV_NET model, I added each user's in-degree, out-degree, betweenness centrality measure, and total negative influence. The ADV_NET model requires explicit construction of the network to compute the betweenness centrality measure and the total negative influence. As a result, it may be time-consuming and computationally intensive to compute these for largeer networks. However, if these two features show a considerable increase in the model's performance, then it may be worthwhile for Verizon to pursue further work in using these features for larger networks.
I defined the total negative influence to be the sum of negative influence scores from all incoming edges:

$$TotalNegativeInfluence(u) = \sum_{v \to u} NegInf(v) \tag{5.1}$$

The churn prediction task was then modeled as a classification problem with a binary label indicating whether the user churned in September. The nodes in the network were split into a training set, a validation set, and a testing set. The performance of the three models on the testing set is listed below.

Model	Accuracy	Precision	F1 Score
Verizon Base	0.847	0.821	0.833
SIMPLE_NET	0.884	0.826	0.839
ADV_NET	0.811	0.788	0.753

From these results, we notice the ADV_NET model is outperformed by the base model in all three metrics, but on the other hand, the $SIMPLE\_NET$ model outperforms the base model in all three metrics.

6 Conclusion

From the results of the negative influence modeling, we find that features predict higher negative influence scores primarily correspond to more phone usage. Higher TOT_MOU_PK, SMS_CNT, $DATA\_USG$, and AVG_DUR all correspond to more phone usage. Intuitively, it makes sense that an individual who uses communicates more frequently with others is able to better influence of his peers. The interesting feature is $URBAN\_2$ which indicates that the customer lives in a suburban area. It may be possible that individuals who live in suburban areas are more able to influence their peers compared to urban areas (ie. Suburban neighborhoods are more tightly-knitted communities, so news about a deal by one of our competitors gets around faster), but I cannot rule out that this is merely an artifact of how the network was constructed.

With regards to churn prediction, I found that incorporating the negative influence scores and betweenness centrality did not improve on Verizon's current models. However, adding the in-degree and the out-degree as features appears to marginally improve the performance of the churn model. This may be due to conformity effects in that an individual who communicates primarily with other Verizon customers may be less inclined to churn to avoid having a different wireless carrier compared to his peers.

One path for future work is identifying how interacting with individuals who are not Verizon customers affects the likelihood that a customer churns. Non-Verizon customers were excluded during the construction of the network, but we might expect that interacting with people who are subscribed to our competitors may influence our customers to switch to match their peer's wireless carrier.

7 Work Cited

1. S. Motahari, T. Jung, H.Zang. Predicting the Influencers on Wireless Subscriber Churn. IEEE WCNC 2014.

2. P. Kusuma, D. Radosavljevik, F.W. Takes, P. Putten. Combining Customer Attribute and Social Network Mining for Prepaid Mobile Churn Prediction. 22nd Belgium Dutch Conference on Machine Learning. 2013.